Phonological Externalization volume 4 (2019), 93-103. Hisao Tokizaki (ed.), Sapporo University.

The Perception of L2 Information Focus Marking

Atsushi Fujimori[1], Noriko Yamane[2], Mineharu Nakayama[3],
Noriko Yoshimura[1] and Ian Wilson[4]
University of Shizuoka[1], Hiroshima University[2], The Ohio State University[3]
and University of Aizu[4]

ABSTRACT. This study examined the perception of English information focus by Japanese EFL learners who encounter difficulties in producing sentence-level prosody (Fujimori, Yoshimura and Shirahata 2014). In a perception task, high-intermediate Japanese EFL learners could correctly perceive only early narrow focus, which is attributed to the Japanese-specific prosodic property of downstep contours.[*]

Keywords: L2 English, prosody, information focus, perception, L1 transfer

1. Introduction

The current study investigates how native speakers of English and Japanese-speaking learners of English perceive English prosody which interfaces with information focus. Acquisition at interfaces between syntax, information structure and prosody is challenging for L2 learners and has been one of the major concerns in L2 study (Sorace 2011, Nakayama and Yoshimura 2015), particularly if the L2 differs substantially from the L1.

This paper is organized as follows: Section 2 outlines the linguistic background of information focus in English and Japanese, followed by a review of previous studies in L2 English. Section 3 explains the experimental procedure. Sections 4 and 5 provide the results and discussion. Section 6 concludes the paper.

2. Background

2.1 Information Focus

Information focus is a non-presupposed part of the sentence and typically observed in an answer to the *wh* constituent in a *wh*-question. English information focus is marked prosodically with a pitch prominence (Beckman 1986) and its placement is flexible and

[*] This study has been partially supported by the Japan Society for the Promotion of Science Grant-in-Aid for Scientific Research (C) 17K02969. Their support is gratefully appreciated. We thank the audience of PhEx 7 for the fruitful discussion. We also thank the participants and Eric Lenhardt in helping us collect data.

context-dependent (Zubizarreta 1998).[1] Transitive constructions allow three different focus types with distinct pitch patterns, as in (1) through (3) (Jannedy 2002).

(1) Q: Who will marry Marrianna?
A: [JEremy]$_{ENF}$ will marry Marriana. (early narrow focus, ENF)

(2) Q: Who will Jeremy marry?
A: Jeremy will marry [MarriAnna]$_{LNF}$. (late narrow focus, LNF)

(3) Q: What will happen?
A: [JEremy will marry MarriAnna]$_{BRF}$. (broad focus, BRF)

The first focus type places a nuclear accent on the subject with an immediate steep pitch fall (i.e., post-focus compression) and the rest of the sentence remaining low, as in Figure 1 (Early Narrow Focus, ENF).

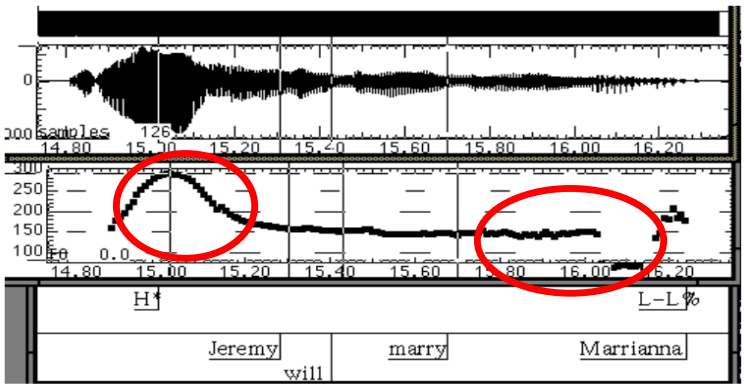

Figure 1: Pitch pattern of ENF
(cited from Jannedy 2002: 37, Figure 3.2 with modification)

The second focus type places two separate nuclear accents on the subject and object with a post-focus compression after the object, as in Figure 2 (Late Narrow Focus, LNF).

[1] English information focus is more prominent than contrastive focus, but only with pitch, not with intensity and duration (Beckman 1986, Breen et al. 2010).

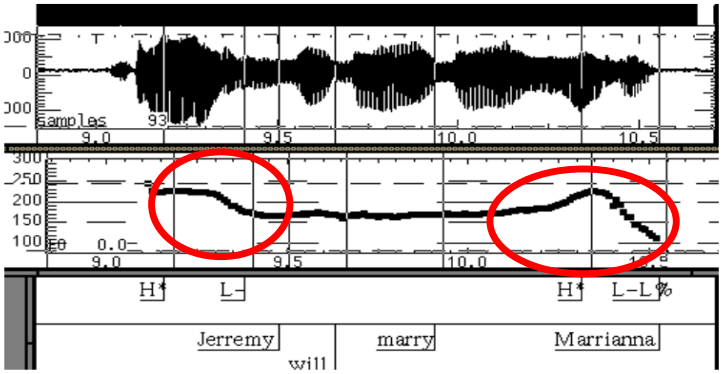

Figure 2: Pitch pattern of LNF

(cited from Jannedy 2002: 38, Figure 3.3 with modification)

The third focus type targets the entire sentence for high pitch with no post-focus compression after the object, as in Figure 3 (Broad Focus, BRF).

Figure 3: Pitch pattern of BRF

(cited from Jannedy 2002: 36, Figure 3.1 with modification)

Japanese information focus can be encoded morphologically with the marker -*GA* (4A) (Kuno 1973). In (4), the answer supplies new information as a response to *dare* 'who' in the previous question. Taro is thus information focus with the marker -*GA*.

(4) Q: (Kyodai-no naka de) dare-ga dokushin desu ka
brother-GEN among who-NOM single is Q
'(Among your brothers), who is single?'

A: [Taroo]$_F$ -GA dokushin desu.
Taro-FOC single is
'Taro is single.'

The sentence-initial word also receives a pitch prominence which initiates downstep contours (Pierrehumbert and Beckman 1988). Therefore, Japanese information focus is often encoded morphologically and phonologically at the left edge of the sentence.

2.2 Prosody in L2 English

Beckman (1986) examined the perception of lexical prosody by Japanese EFL learners (see also Sugahara 2016). In perceiving English lexical stress of disyllabic noun-verb pairs (e.g., *DIgest* (noun) vs *diGEST* (verb)), Japanese EFL learners were more sensitive to pitch than to duration and intensity while native speakers of English equally relied on multiple prosodic cues including duration. She concluded that the EFL learners' performance was attributed to L1 transfer as Japanese is a pitch-accent language.

Aoyama and Guion (2007) investigated how adult Japanese EFL learners would produce lexical prosody in reading aloud sentences such as (5) and (6). Compared to native speakers of English, the Japanese EFL learners more heavily relied on pitch, but not duration, in producing content words than native speakers of English, as in Table 1.

(5) I'm fine.
(6) They went to school.

Table 1: Average duration ratios and pitch ranges

	Duration ratio to the entire sentence		Pitch range (semitone)	
Group	Content words (e.g., *fine, went, school*)	Functional words (e.g., *I'm, they, to*)	Content words	Functional words
NSE	.47 (SD .01)	.20 (.02)	2.7 (1.1)	1.5 (.7)
JEFL	.48 (.02)	.23 (.07)	3.9 (.7)	1.5 (.8)

(cited from Aoyama and Guion 2007: 293, Table 6 with modification)

These studies have shown that Japanese EFL learners are sensitive to pitch at the lexical level of L2 English. One might wonder whether Japanese EFL learners can produce and perceive sentence-level prosody pertaining to information focus. Fujimori, Yoshimura and Shirahata (2014) reported results of a read-aloud task with 10 low-level Japanese EFL learners and eight native speakers of English. The native speakers of English placed the highest pitch on the focused word in question-answer congruence, as in (7) and Figure 4.

(7) Q: Where did you go last Sunday?
A: I went [FIshing]$_F$ with my friend in the river.

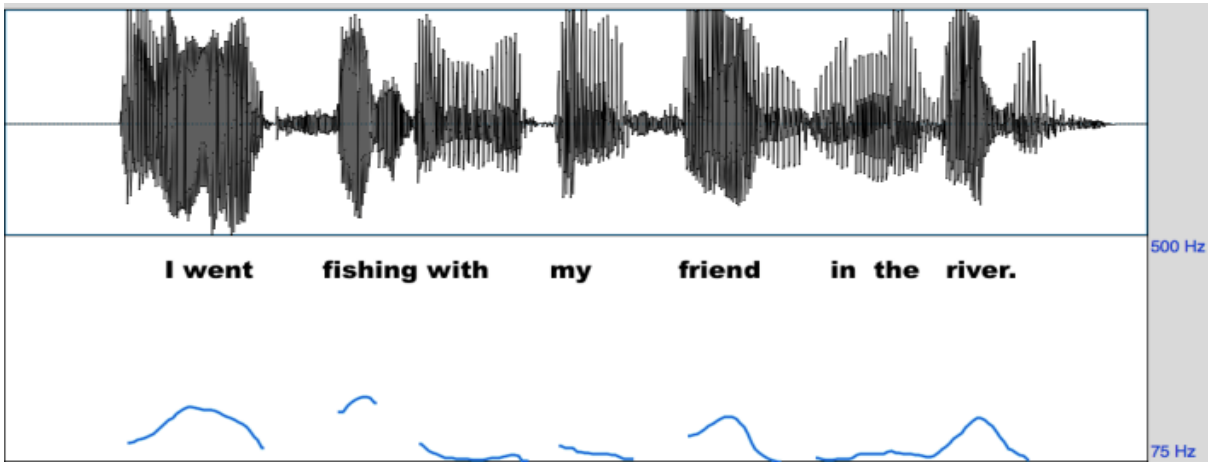

Figure 4: Appropriate prosodic prominence by a native speaker of English
(cited from Fujimori et al. 2014: 51, Figure 1 with modification)

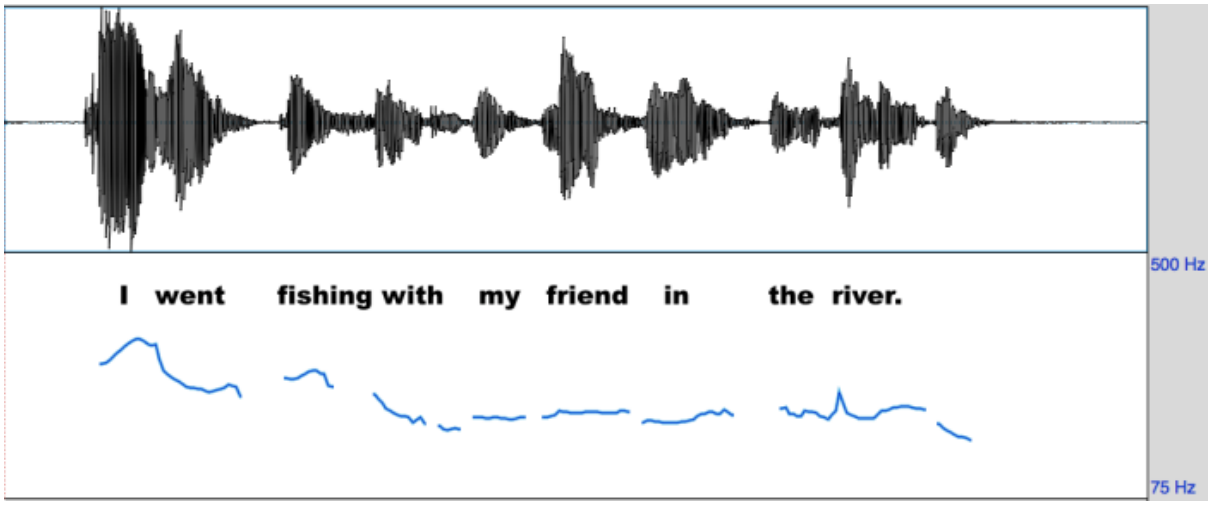

Figure 5: Typical downstep contours by Japanese EFL learners
(cited from Fujimori et al. 2014: 52, Figure 2 with modification)

In contrast, the Japanese EFL learners placed the highest pitch on the subject, as in Figure 5. Their production is affected by the Japanese-specific property of downstep contours. Such sentence-level prosodic L1 transfer effects are reported by Nava (2008). Nava conducted a read-aloud task with Spanish-speaking learners of English. Spanish is a right-edge prominence language and information focus rigidly appears at the sentence-final position, as in (8).

(8) a. ¿De qué te ríes?
at what you laugh-PRS-PROG
'What are you laughing at?'

b. ¡Un pingüíno está [bailando]$_F$!
a penguin be-PRS-3SG dance-PROG
'A penguin is dancing.'

The intermediate to high-level L1 Spanish-L2 English speakers incorrectly placed a pitch prominence on the sentence-final position, as in (9b).

(9) a. Why are you looking out the window?
b. Madonna just walked [by]$_F$! (L1 Spanish-L2 English)
c. [Madonna]$_F$ just walked by! (L1 English)

Japanese EFL learners encounter difficulties in producing sentence-level prosody, due to the discrepancy between L1 Japanese and L2 English in the placement of pitch prominence. The remaining problem is whether Japanese EFL learners can perceive sentence-level prosody pertaining to information focus. If sentence-level prosodic L1 transfer occurs, it is predicted that Japanese EFL learners will correctly perceive only sentence-initial foci, that is, ENF.

3. Experimental Procedure

To test the prediction, a perception task was conducted with nine native speakers of English (NSEs) and 45 Japanese EFL learners (JEFLs) with CEFR B1 or B2 level (average TOEIC 714.33, SD 77.21). In the task, the participants were first asked to read a particular context displayed on a PC monitor. After a 15 second delay, the participants listened to a context-relevant *wh*-question to elicit either ENF, LNF, or BRF. Then they were asked to select the best answer from a set of three potential recorded answers, which differed only in prosody. A sample token is shown in (10) where the question and the answer choices in red were not displayed during the session. The question-answer set was repeated twice. Three tokens were provided for each focus type in a Latin square design, i.e., each participant encountered nine test tokens during the session.

(10)

Context: Mary and Jennifer went shopping at a department store last weekend. Mary bought boots for her mother's birthday.

Question 1: Who bought boots for her mother?

A) Mary bought [BOOTS]$_{LNF}$ for her mother.

B) [MAry]$_{ENF}$ bought boots for her mother.

C) [MAry bought BOOTS for her MOther]$_{BRF}$.

D) I don't know.

The test tokens were recorded with a male native speaker of North American English in a sound-attenuated room. He was asked to read aloud each question-answer pair with appropriate intonation. A question was later combined with a set of three different answer choices in audio software Audacity. Each answer included an adjunct PP (e.g., "for her mother") as well as a transitive verb, which gave listeners more cues for prosodic discrimination.

4. Results

The NSEs' overall correct response rates were unexpectedly low: 81.5% for ENF, 81.5% for LNF, and 66.7% for BRF. The correct response rates in the LNF and ENF conditions were slightly higher than that in the BRF.

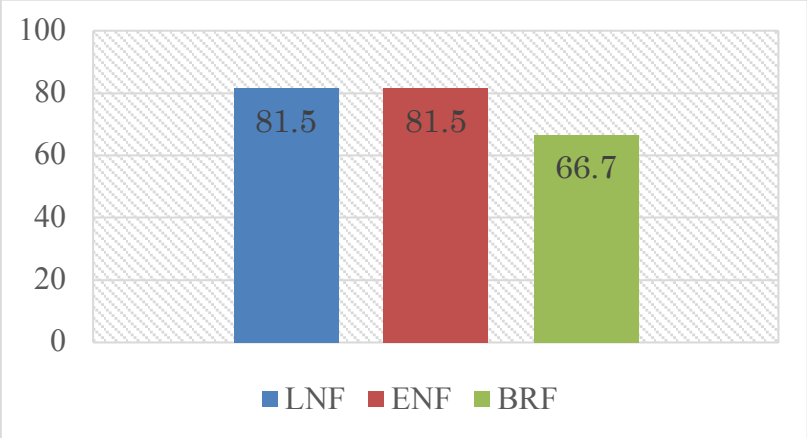

Figure 6: NSEs' correct response rates (%)

There were three major error types identified in the analysis of NSEs' incorrect answers, as shown in Table 2: i) choosing answer A (LNF) for type C (BRF) question, ii) choosing answer C (BRF) for type A (LNF) question, and iii) choosing answer C (BRF) for type B (ENF) question.

Table 2: Three major error types in NSEs' responses

Error type	Token	Errors	Participants	Prosodic properties of answer choices
i) Choosing answer A for type C question	T1	2	P5, P8	Answers A and C are similar in pitch pattern.
	T4	2	P4, P5	
	T5	1	P7	
	T6	1	P6	
	Subtotal	6		
ii) Choosing answer C for type A question	T4	1	P9	Answers A and C are similar in pitch pattern.
	T5	2	P4, P8	
	T6	2	P2, P7	
	Subtotal	5		
iii) Choosing answer C for type B question	T4	1	P8	Answers B and C are similar in pitch pattern.
	T5	1	P9	Answers B and C are partially similar in pitch pattern. Also pitch rise on PP in answer C, but not in B.
	T6	1	P8	
	T7	1	P4	
	Subtotal	4		
iv) Choosing answer B for type C question	T2	1	P3	Answers B and C are partially similar in pitch pattern. Also pitch rise on PP in answer B, but not in C.
	Subtotal	1		

In the measurement of prosodic cues, the answer choices A, B, and C showed similar pitch contours in Tokens 4 to 6 while they did differ in duration. Taking Token 6 for example, blue curves in Figure 7 show similar downstep contours for answer choices A (LNF) and C (BRF) while the duration of the object in answer choice A was 386 ms and longer than that of answer choice C (296ms). Because of the similarity in pitch contours, several NSE participants made quite a few errors in perceiving those tokens.

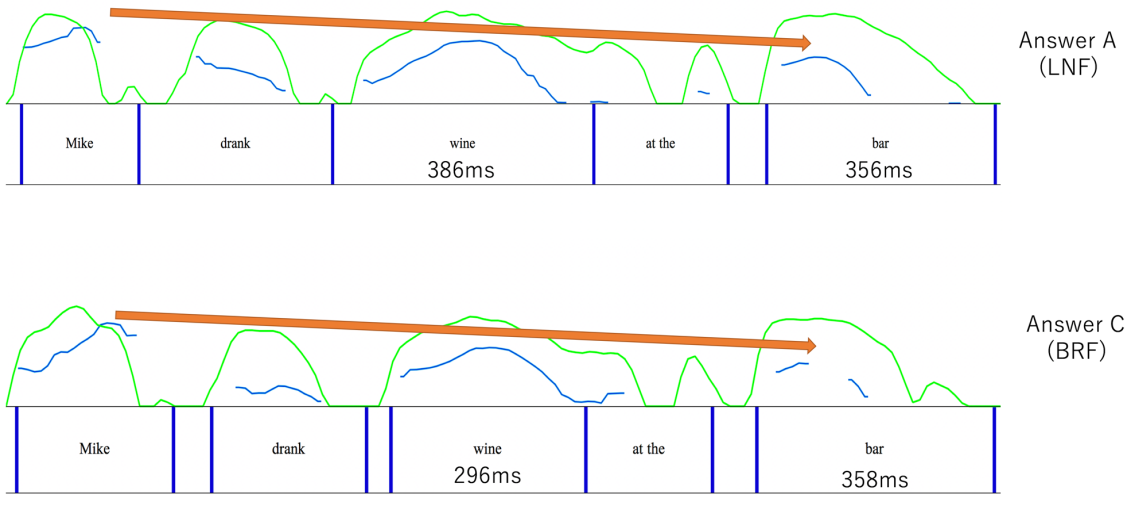

Figure 7: Pitch patterns and relative duration of OBJ to within-PP noun

These facts suggest that i) there was a variety of prosodic realization of information focus within a native speaker of English and that ii) some of the NSEs primarily relied on pitch in

perceiving the prosody of information focus while others relied on duration. These have never been pointed out in the previous literature (cf. Beckman 1986, Breen et al. 2010).

The three tokens were removed from the current analysis and in the rest of the six test tokens the NSEs performed well: The correct response rates were 100% for LNF, 88.8% for ENF and 72.2% for BRF. Clear-cut pitch pattern differences were also identified between the three answer choices A, B, and C.

The NSEs' correct response rates for the six tokens were compared with the JEFLs' counterparts. An ANOVA (2 groups x 3 focus types) revealed that there were main effects on groups (F (1, 106) = 443.75, $p < .01$) and focus types (F (2, 212) = 6.67, $p < .01$). While there were no interactions (F (2, 212) = .65, p = .53), post-hoc tests showed that the control performed significantly better than the L2 learners in LNF and BRF, but not in ENF, as shown in Table 3. There was also a significant type difference between BRF vs ENF and LNF in L2 English.

Table 3: Response rates by groups and by types (%)

QUESTION TYPE	RESPONSE (L1 English)			RESPONSE (L2 English)		
	ENF	LNF	BRF	ENF	LNF	BRF
ENF	88.8	5.6	5.6	75.3	13.5	11.2
LNF	0	100	0	8.9	67.8	23.3
BRF	5.6	22.2	72.2	11.1	50.0	38.9

5. Discussion

The current findings support the prediction: The high-intermediate JEFL learners demonstrated native-like perception of only sentence-initial foci, due to positive L1 transfer. While their correct response rates for LNF and BRF were significantly lower than those of the NSEs, there was also a significant discrepancy between LNF and BRF for the JEFL learners. One might say that these facts are attributed to the crosslinguistically attested Complement Law (Nespor et al. 2008) which assigns a default focus marker to the verb complement. In fact, nine JEFL learners constantly chose LNF, instead of BRF. So, it is expected that low-level JEFL learners would choose more LNF type answers. However, the findings could also be accounted for in terms of perceptual salience as well. As seen in the previous section, a major prosodic cue to differentiate between LNF and BRF is a post-focus compression after the object. Then it might have been the case that the L2 learners were insensitive to the subtle but distinctive pitch contour and that their responses to the BRF type questions split between BRF

and LNF. To identify the cause of the L2 difficulty, therefore, we further need to test L2 learners' perception in a variety of sentences containing double object constructions and prepositional dative constructions, as in (11).

(11) a. John gave Mary a book.
b. John gave a book to Mary.

In these constructions indirect objects and prepositional datives are internal arguments while the prepositional phrases used in the current experiment are adjunct. If the cause of the L2 difficulty is attributed to prosodic rules interfacing with syntax, it is predicted that L2 learners will show asymmetries between argument and adjunct prepositional phrases in perceiving focus prosody. This prediction needs to be tested in future research.

6. Conclusion

The current study has shown that high-intermediate Japanese EFL learners could correctly perceive only sentence-initial foci in L2 English, due to L1 transfer. The finding leads to a question of whether far more advanced Japanese EFL learners can ultimately perceive late narrow focus and broad focus. Without perceiving the three focus types, Japanese EFL learners are expected not to be able to produce them. Another question is whether and how their ability of perceiving sentence-level prosody in L2 English can be trained. It is worthwhile examining whether instructing prosodic prominences on foci or post-focus compression is effective to train L2 learners' perceptual ability.

References

Aoyama, K. and S. G. Guion. (2007). Prosody in second language acquisition: Acoustic analysis of duration and F0 range. In O.-S. Bohn & M. J. Munro (eds.), *Language experience in second language speech learning. In honor of James Emil Flege* (pp. 281-297). Amsterdam/Philadelphia: John Benjamins.

Beckman, M. E. (1986). *Stress and Non-Stress Accent*. Netherlands Phonetic Archives Series No. 7. Dordrecht: Foris.

Breen, M., E. Fedorenko, M. Wagner, and E. Gibson. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes* 25 (7/8/9), 1044-1098.

Fujimori, A. N. Yoshimura, and T. Shirahata. (2014). Acquisition of prosodic focus marking by Japanese ESL learners. *IEICE Technical Report* 114 (176), 49-53.

Jannedy, S. (2002). *Hat patterns and double peaks: The phonetics and psycholinguistics of*

broad versus late narrow versus double focus intonations. Doctoral dissertation, OSU.

Kuno, S. (1973). *The Structure of the Japanese Language*. Cambridge, MA: MIT Press.

Nakayama, M. and N. Yoshimura. (2015). The modularity of grammar in L2 acquisition. In M. Nakayama (ed.), *Handbook of Japanese Psycholinguistics* (pp. 235-270). Berlin / Boston: De Gruyter Mouton.

Nava, E. (2008). Prosody in L2 Acquisition. In R. Slabakova, J. Rothman, P. Kempchinsky & E. Gavruseva (eds.) *Proceedings of GASLA 2007*, 155-164.

Nespor, M., M. Shukla, R. van de Vijver, C. Avesani, H. Schraudolf, and C. Donati. (2008). Different phrasal prominence realizations in VO and OV languages. *Lingue e Linguaggio* 7, 1-29.

Pierrehumbert, J. and M. Beckman. (1988). *Japanese Tone Structure*. Cambridge, MA: MIT Press.

Sorace, A. (2011). Pinning down the concept of 'interface' in bilingualism. *Linguistic Approaches to Bilingualism* 1, 1-33.

Sugahara, M. (2016). Is Japanese listeners' perception of English stress influenced by the antepenultimate accent in Japanese? Comparison with English and Korean listeners. *Doshisha Studies in English* 96, 61-111.

Zubizarreta, M. (1998). *Prosody, Focus, and Word Order*. Cambridge, MA: MIT Press.